ÖREBRO UNIVERSITET
AKADEMIN FÖR
NATURVETENSKAP OCH TEKNIK

Lösningar till tentamen i

Numeriska metoder för civilingenjörer

DT508G

2018-??-??

Observera att lösningsförslag inte är en fullständig lösning.

1. Lös följande delproblem. [10p]

   (a) Skriv om polynomet $p(x) = 4x^3 - 2x^2 + 3x - 4$ på på nästlad form (nested form). [1p]

   (b) Generalisera resultatet i (a) till ett generellt polynom av grad n. [3p]

   (c) Härled antal flyttalsberäkningar (multiplikation, addition, subtraktion, division räknas som en flyttalsberäkning) för den algoritm ni tagit fram i (b). [3p]

   (d) Vilken typ av interpolation ger koefficienter som direkt kan användas i algoritmen ni tagit fram i (b)? [1p]

   (e) Antag att vi har polynomet $p(x) = 1 - x + x^2 - x^3 + x^4 - x^5 + ... - x^{99}$ som ska beräknas numeriskt för x mycket nära 1. Kalle väljer att beräkna polynomet genom följande loop: [2p]

   ```
   s = 0;for i=2:2:100,s = x^(i-2)-x^(i-1)+s;end
   ```

   men Stina väljer att beräkna med formeln $(1 - x^{100})/(x + 1)$. Förklara varför Stinas formel ger bättre noggrannhet.

   Lösning:

   (a) Man får $p(x) = -4 + x(3 + x(-2 + x(4)))$.

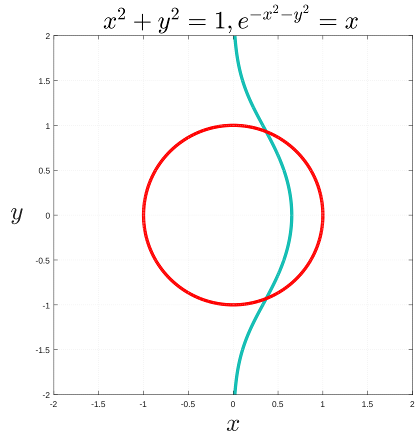
   (b) Se boken kapitel 0, Fundamentals.

   (c) Komplexiteten är n multiplikationer och $2n$ additioner.

   (d) Newtoninterpolation eftersom de koefficienter man får där direkt kan användas för nästlad form.

   (e) I Kalles variant blir det kancellation i varje steg i for-loopen men i Stinas formel blir det endast en kancellation. Notera att dessutom blir kancellationseffekten värre i Kalles variant eftersom termerna är mer lika i storleksordning.

2. Event AB planerar en ny berg och dalbana och funderar på en kurvform av typen kurvan $f(x) = x\cos(x) - \sin(x)e^x$, $x \in [-12, 2]$ se figur nedan. [10p]



Konstruktionen av vagnarna kräver att dom bestämmer precis när vagnen passerar $f(x) = 0$. Konsultföretaget som anlitas för detta kräver då att ni löser följande delproblem.

(a) Beskriv i detalj en metod som med bra noggrannhet identifierar startlösningar för Newton-Raphsons metod för alla rötter i intervallet. [2p]

(b) Beskriv i detalj Newton-Raphsons metod för detta problem där ni även ska skissa på algoritm, ange startvärde, konvergenskriterium (kan vara flera) och uppskattning av noggrannheten i er numeriska lösning. [6p]

(c) Ett nollställe ligger uppenbarligen i $x = 0$. Förklara varför Newton-Raphsons metod konvergerar långsammare för denna rot. Vilken konvergenshastighet kan man förvänta sig i denna rot? [2p]

Lösning:

(a) Se intervallhalvering i boken.

(b) Se boken och er egen Matlabkod. Startvärde tar ni från intervallhalvering och konvergenskriteria är att iterera tills $|x_{k+1} - x_k| \leq tolx$ och $|y_k| \leq toly$. Uppskattning av noggrannhet i lösningen ges av konvergenskriterierna.

(c) Derivatan är $f'(x) = \cos(x) - e^x - x\sin(x)$ och $f'(0) = 0$ samt andraderivatan är $f''(x) = e^x - 2\sin(x) - x\cos(x)$ med $f''(0) = -1$ vilket gör att Newton-Raphson konvergerar linjärt.

3. Avgör om följande påståenden är sanna eller falska och ge en kort motivering till ert svar. [10p]

(a) Det uppstår alltid kancellation när man numeriskt beräknar derivatan av en deriverbar funktion. [2p]

(b) Jacobis metod konvergerar kvadratiskt för ickelinjära ekvationer. [2p]

(c) Extrapolation kan användas för att förbättra lösningen vid numerisk differentiering och integration. [2p]

(d) Explicit Euler fungerar bättre än implicit Euler för styva problem. [2p]

(e) Konditionstalet för linjära ekvationssystem $Ax = b$ är en övre uppskattning av kvoten mellan relativa felet i lösningen och relativa felet i indata (dvs relativa felet i A och b). [2p]

Lösning:

(a) Ja, eftersom man alltid subtraherar två små tal och dividerar dessutom med ett litet tal.

(b) Nej, Jacobis metod används för att lösa linjära ekvationssystem.

(c) Ja, eftersom man känner felutvecklingen kan man få en högre ordningens approximation genom att beräkna för olika h.

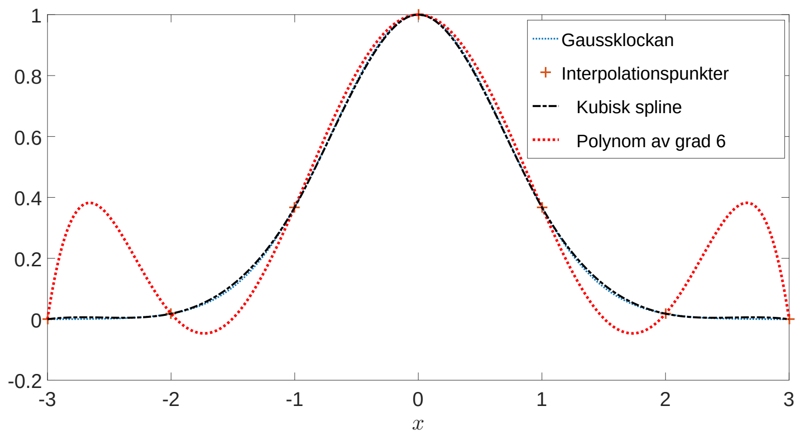
(d) Nej, tvärt om.

(e) Ja, det är defintionen av konditionstalet som man kan beräkna som $\kappa = \|A\|\|A^{-1}\|$ där man oftast inte beräknar inversen utan approximerar konditionstalet med andra mer effektiva metoder.

4. Herr Interpolation är en stor beundrare av Gaussklockor så han vill approximera en sådan med hjälp av interpolation. För enkelhetens skull börjar han med Gaussklockan [10p]

$$f(x) = e^{-x^2}, x \in [-3, 3]$$

se figur nedan. Eftersom han är snål använder han väldigt få interpolationspunkter. Förklara för Herr Interpolation varför spline-interpolanten blir så mycket bättre än högre ordningens polynom genom att lösa följande uppgifter.



(a) Vad kallas det fenomen som kan uppkomma vid interpolation med högre ordningens polynom och varför uppkommer detsamma? [2p]

(b) Beskriv så detaljerat ni kan hur fenomenet i (a) vid interpolation med polynom (ej styckvisa polynom) kan undvikas. [3p]

(c) Beskriv vad som menas med en kubisk spline och varför den är så mycket bättre i detta fall. [3p]

(d) Antag att Herr Interpolation gör misstaget att inte beräkna de givna interpolationspunkterna exakt utan med ett litet fel av storleksordning, säg, 10^{-4}. Resonera hur ni då kan finna en approximation av Gaussklockan. Ni har full frihet att betrakta olika relevanta problemställningar och vad dessa då skulle ge för resultat i fråga om noggrannhet, effektivitet och användbarhet. [2p]

Lösning:

(a) Det kallas Runges fenomen och uppkommer på grund av att feltermen vid interpolation kan bli stor för vissa val av interpolationspunkter. I detta fall är punkterna ekvidistanta vilket inte är optimalt.

(b) Här fungerar Chebyshevpunkter utmärkt som minimerar feltermen. De väljs som projektioner av punkter på enhetscirkeln som sedan skalas för att stämma överens med det givna intervallet.

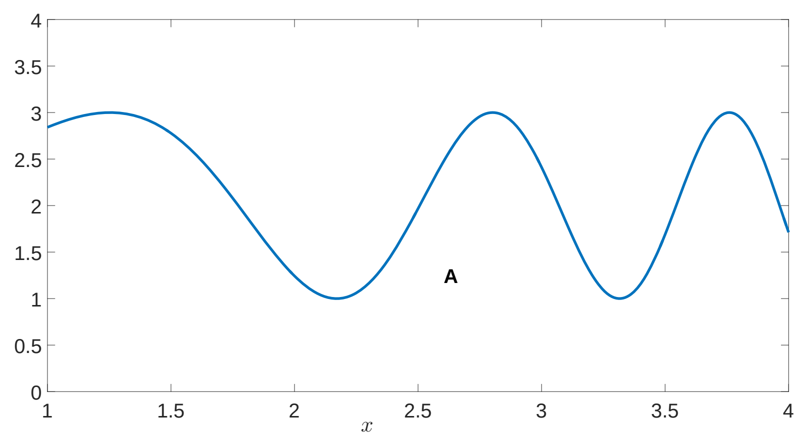
(c) En kubisk spline är ett styckvis polynom där varje delintervall är en kubisk funktion. Den interpolerar alla punkter och dessutom är derivatan och andraderivatan kontinuerlig i de inre punkterna. Eftersom den naturligt undviker Runges fenomen och dessutom anpassar sig till hur kurva ser ut på respektive intervall ger den ett mycket bra resultat. Det kommer dock till en något högre kostnad som dock är försumbar för detta lilla problem.

(d) Man kan använda minstakvadratmetoden dvs anpassa ett polynom till ge givna punkterna. T.ex. om man tar ett 4'e gradspolunom får man ett minstakvadratproblem med en matris som har 7 rader (data) och 5 kolumner (obekanta koefficienter). Minstakvadratproblemet kan sedan lösas med QR-faktorisering med ett beräkningsfel betydligt mindre än felet i de givna punkterna. Ett bättre sätt är att använda en approximerande spline vilket dock är betydligt svårare att implementera i Matlab (finns färdiga funktioner för detta dock).

5. Herr Integral tänker bygga en ny golfbana. En del av banan kräver dyr markberedning och därför vill Herr Integral uppskatta arean under funktionen [10p]

$$y(x) = 2 + \sin(x^2), x \in [1,\ 4]$$

se figur nedan.



Inför en indelning i x genom $x_k = 1 + 3(k-1)/N, k = 1, \ldots, N+1$.

(a) Beräkna $h = x_{k+1} - x_k$ och beräkna arean under den linjära funktionen som går genom $y(x_k)$ och $y(x_{k+1})$. Kalla den arean för A_k. [2p]

(b) Formulera Trapetsmetoden genom att bilda summan $A_h = \sum_k A_k$ där A_k är från (a). Ange undre och övre summationsgräns. [2p]

(c) Trunkeringsfelet i Trapetsmetoden för detta problem är [4p]

$$\frac{3}{12}h^2y''(c), c \in [1, 4]$$

Hjälp Berra att ge en övre gräns på Trunkeringsfelet för hans problem då $N = 300$.

(d) Om Berra inte är nöjd med felet i (c) vad kan han göra för att förbättra detta? [2p]

Lösning:

(a) Vi har $A_k = h(y(x_k) + y(x_{k+1}))/2$, $h = 3/N$.

(b) Trapetsmetoden blir $\sum_{k=1}^{N} A_k$.

(c) Eftersom $y'' = 2\cos(x^2) - 4x^2\sin(x^2)$ får vi en övre gräns som

$$|\frac{3}{12}h^2y''(c)| \leq \frac{3}{12}0.01^2 \cdot 66 = 0.0017.$$

(d) Det enklaste är att minska h. Berra kan också införa en adaptiv grid som har en tätare grid där y varierar mer dvs i övre delen av intervallet. Extrapolation är ytterligare en effektiv metod att få ned felet.

6. Antag att $A \in \mathbb{R}^{n\times n}, b \in \mathbb{R}^n$ och att ekvationssystemet $Ax = b, x \in \mathbb{R}^n$ har en entydig lösning. Besvara följande delfrågor. [10p]

(a) Ge ett exempel på ett ekvationsystem där $n = 2$ som har entydig lösning men där Gausselimination inte fungerar. [2p]

(b) Nedanstående kod i Matlab gör Gausselimination. Gör de nödvändiga ändringarna i koden för att generera L-matrisen i en LU-faktorisering av A. [2p]

```
for j=1:(n-1)
   for i = j+1:n
       mult = A(i,j)/A(j,j); count = count+1;
       for k=j+1:n
           A(i,k)= A(i,k) - mult*A(j,k); count = count+2;
       end
       b(i)= b(i) - mult*b(j); count = count+2;
   end
end
```

(c) Använd LU-faktoriseringen till A för att algebraiskt visa hur man i två steg med framåt- respektive bakåtsubstitution beräknar lösningen till ekvationssystemet. [2p]

(d) Varför kan det i vissa fall vara väldigt effektivt att använda en LU-faktorisering? [1p]

(e) Härled komplexiteten för att lösa $Ly = b$ där L är en undertriangulär matris. [3p]

Lösning:

(a) Tag t.ex. [2p]

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Som har lösningen $x_1 = x_2 = 1$ men där Gausselimination ger pivotelementet lika med noll. Det krävs rad- eller kolumnpivotering.

(b) Lägg till `L(i,j)= mult;` på rad 3. [2p]

(c) Lös först $Ly = b$ med framåtsubstitution sedan $Ux = y$ med bakåtsubstitution. [2p]

(d) Om man vill lösa $Ax_i = b_i, i = 1, \ldots, N$ då N är mycket större än 1. [1p]

(e) Se boken. [3p]